AN APPROACH TO SIMULTANEOUS TOLERANCE INTERVALS IN REGRESSION

By A. L. Wilson

Biometrics Research Services, Canada Department of Forestry and Rural Development

1. Introduction. Lieberman and Miller (1963) extended the tolerance-interval theory of Wald and Wolfowitz (1946) and Wallis (1951) so as to permit us to put a tolerance band about a regression line. Given any pair (δ, γ) of probabilities, the band can be constructed in such a way that the probability will be approximately δ of obtaining a line whose tolerance band includes a proportion γ of the population. Adoption in this paper of a different approach results in some departures from familiar phenomena:

(1) The way we begin by explicit construction of the set on which the tolerance statements are to hold strengthens our intuition regarding the procedure, how far it falls short of optimality, and (to some extent) the nature of the Lieberman-Miller set.

(2) Because our procedure yields a tolerance band wider, at $x = \bar{x}$, than the L-M variety, but narrower for larger values of $|x - \bar{x}|$, it may be anticipated that it will have advantages for extrapolation outside the observed range of x.

(3) A separate calculation is necessary at each value of x for which the bandwidth is needed. However, one has only to solve a quadratic equation, and once the work has been organized for desk calculator, five to ten minutes suffices for each value.

2. The problem. Let $x_1, \cdots, x_n$ be given, and for convenience let $x_1 + \cdots + x_n = 0$ and $x_1^2 + \cdots + x_n^2 = 1$. The components of $\mathbf{y} = (y_1, \cdots, y_n)'$ are assumed independently normal with $E(y_i) = \alpha + \beta x_i$ and $\operatorname{Var}(y_i) = \sigma^2$. The regression equation will be $\hat{y}(x) = \bar{y} + bx$, where $b = \sum(xy)$; and $s^2 = (\sum (y^2) - n\bar{y}^2 - b^2)/(n-2)$. It will be obvious how the treatment in this paper extends to the multiple-regression model.

Let $f(y; \alpha, \beta, \sigma)$ be the density of a variable whose distribution is $N(\alpha + \beta x_0, \sigma^2)$. Given probabilities δ and γ, and x_0, a value of the transformed x, Wallis (1951) showed how a constant $\nu = \nu(x_0) > 0$ could be determined such that if we define, in the n-dimensional space of $\mathbf{y}$ or the 3-dimensional space of $(\bar{y}, b, s)$, the set

$$R(x_0) = \{\textstyle\int_{\bar{y}+bx_0-\nu s}^{\bar{y}+bx_0+\nu s} f(t; \alpha, \beta, \sigma)\, dt \geqq \gamma\},$$

then $P(R(x_0)) = \delta$ approximately.

In the present paper we define a set T of probability approximately δ, then derive a relationship $\lambda = \lambda(x_0)$ such that if $S(x_0) = \{\int_{\bar{y}+bx_0-\lambda s}^{\bar{y}+bx_0+\lambda s} f(t; \alpha, \beta, \sigma)\, dt \geqq \gamma\}$,

Received 9 January 1967; revised 15 May 1967.

$T \subset S(x_0)$ for all x_0. Naturally, we make each set $S(x_0)$—which is to say each $\lambda(x_0)$—as small as we can.

3. Preliminaries. Not only does $\hat{y} \pm \lambda s$ suggest itself as the most "natural" form for tolerance limits; it has a valuable invariance property. Let $f(y; \alpha, \beta, \sigma)$ represent the density of $y \sim N(\alpha + \beta x, \sigma^2)$. If

$$\int_{\bar{y}+bx-\lambda s}^{\bar{y}+bx+\lambda s} f(t; \alpha, \beta, \sigma)\, dt = \gamma,$$

and we make the transformation $z = (t - \alpha - \beta x)/\sigma$ so as to have $\int f(t; 0, 0, 1)\, dt$, the limits of integration become $(\bar{y} - \alpha)/\sigma + [(b - \beta)/\sigma]x \pm \lambda(s/\sigma)$. Consequently, in determining λ, we can assume that the transformation has already been made, that $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I})$, $\bar{y} \sim N(0, n^{-1})$, $b \sim N(0, 1)$ and $(n - 2)s^2 \sim \chi^2_{n-2}$. The independence of $\bar{y}$, b and s is important for this section.

To make calculation feasible, we are going to use the familiar approximation $(2\chi^2_{n-2})^{\frac{1}{2}} \sim N((2(n - 2) - 1)^{\frac{1}{2}}, 1)$. When a line is fitted on the basis of, say, ten or fifteen observations, this approximation will be somewhat inexact. But it is conjectured that when, in the following paragraph, we construct a new χ^2 statistic, a proportion of the imprecision will be "ironed out." And in the next section it will be argued that departures from ostensible probability levels can exist without affecting the reliability of the eventual result.

According to the approximation above, which we employ from now on as if it were exact, $s \sim N(k, 1/(2n - 4))$, where $k^2 = (2n - 5)/(2n - 4)$. If

$$T = \{n\bar{y}^2 + b^2 + (2n - 4)(s - k)^2 \leqq 2c^2\}, \tag{1}$$

constituting the interior of the ellipsoid of concentration for the variables $(\bar{y}, b, s)$, $2c^2$ being that value of ${\chi_3}^2$ exceeded with probability $1 - \delta$, then $P(T) = \delta$.

4. Method.

LEMMA. *If* $A = \bar{y} + bx$ *and* $r = \lambda s$, *then*

$$T \subset U(\lambda, x) = \{A^2/(n^{-1} + x^2) + (2n - 4)(r - \lambda k)^2/\lambda^2 \leqq 2c^2\} \tag{2}$$

for all x *and all* $\lambda > 0$.

That is, if $(\bar{y}, b, s)$ is a point of the 3-dimensional ellipsoid, (A, r) will be a point of the related ellipse.

PROOF. Since $n\bar{y}^2 + b^2 = A^2/(n^{-1} + x^2) + (x\bar{y}n^{\frac{1}{2}} - bn^{-\frac{1}{2}})^2/(n^{-1} + x^2) > A^2/(n^{-1} + x^2)$ and $(s - k)^2 = (r - \lambda k)^2/\lambda^2$, the inequality of (1) implies that of (2).

REMARK. Although not setting forth the full argument applicable to a multiple regression surface, we need to be sure that the inequality $n\bar{y}^2 + b^2 > A^2/(n^{-1} + x^2)$ generalizes to that situation. The model is then $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $(n \times p)$ with $\mathbf{X}'\mathbf{X} = \mathbf{I}$. The BLUE of $\boldsymbol{\beta}$ is $\mathbf{b} = \mathbf{X}'\mathbf{y}$. In the case of $\boldsymbol{\beta} = \mathbf{0}$, $\sigma = 1$, the ellipsoid of concentration for the variables $b_1, b_2, \cdots, b_p$, s is $\mathbf{b}'\mathbf{b} + 2(n - p)(s - K)^2 = 2C^2$, where $K^2 = ((2n - p) - 1)/2(n - p))$, and $2C^2$ equals the value of χ^2_{p+1} exceeded with probability $1 - \delta$. If $\mathbf{x}_0$ is an arbitrary $(1 \times p)$ vector $\neq \mathbf{0}$, the tolerance limits for $\hat{y}(\mathbf{x}_0)$ have the form

$\mathbf{x}_0\mathbf{b} \pm \lambda s$. An ellipse of concentration for the variables $A = \mathbf{x}_0\mathbf{b}$ and $r = \lambda s$ is

$$A^2/\mathbf{x}_0\mathbf{x}_0' + 2(n - p)(r - \lambda K)^2/\lambda^2 = 2C^2.$$

It is sufficient to establish that $\mathbf{b}'\mathbf{b} \geqq (\mathbf{x}_0\mathbf{b})^2/\mathbf{x}_0\mathbf{x}_0'$; and this is merely a form of the Schwarz inequality. [The components of $\mathbf{x}_0$ must of course be interpreted as quantities that have been adjusted as was done to make $\mathbf{X}'\mathbf{X} = \mathbf{I}$.]

Consider the relationship between A and r determined by the requirement

$$\int_{A-r}^{A+r} f(t; 0, 0, 1)\, dt = \gamma. \tag{3}$$

For $A = 0$, r has the familiar value such that $P\{-r < y < r \mid y \sim N(0, 1)\} = \gamma$, while as $A \to \pm\infty$, $r - |A| \to r_0$ where $P\{y < r_0\} = \gamma$. The (A, r) curve has asymptotes $r - r_0 = \pm A (r \geqq 0)$, and it was found convenient to represent it by the upper branch of an hyperbola

$$(r - r_0)^2 - A^2 = h^2 \tag{4}$$

(taking the r-axis as the vertical one) h being chosen to achieve a good fit in that part of the curve, found, in the sequel, to be of interest. For each γ, a suitable h can be found by trial and error. In practice it turned out that a very satisfactory value of h was one which minimized the sum of squared deviations $(r - r_0)^2 - A^2 - h^2$ for the eleven values $A = 0(.1)1$ and corresponding r. In the range of interest, the hyperbola (4) lies slightly above the true curve, and it will be seen shortly that this situation guarantees conservative intervals. Table 1 gives approximate values of r_0 and h^2 appropriate to certain selected γ.

It is obvious that if we have a value of λ such that no point of the ellipse (2) lies below the hyperbola (4), $\bar{y} + bx \pm \lambda s$ will be a set of tolerance limits, since

$$T \subset U(\lambda, x) \subset S(\lambda, x) = \left\{\int_{\bar{y}+bx-\lambda s}^{\bar{y}+bx+\lambda s} f(t; 0, 0, 1)\, dt \geqq \gamma\right\}$$

and $P(T) = \delta$. For sufficiently small λ, $U(\lambda, x)$ will intersect the hyperbola; the optimal λ will correspond to tangency of the two conics. When λ has this value, λ' say, designate $U(\lambda', x) = U(x)$; $S(\lambda', x) = S(x)$.

Substituting $A^2 = (r - r_0)^2 - h^2$ in (2), and replacing the inequality sign with equality, we have a quadratic equation in r; the discriminant is itself proportional to a quadratic $\varphi(\lambda)$ in λ. It can be verified that for all values of n, x, δ and γ, $\varphi(\lambda) = 0$ has real roots, of which the larger, λ', yields the desired ellipse $U(\lambda', x) = U(x)$.

TABLE 1

γ	r_0	h^2
.50	0	0.455
.75	0.674	0.250
.80	0.842	0.107
.90	1.28	0.0657
.95	1.65	0.0438
.99	2.33	0.0244

The variable s may be replaced by any statistic proportional to one approximately normally distributed, independent of $\bar{y}$ and b, so long as it has the "invariance" property that its expectation is unaffected by changing $E(\mathbf{y})$, while its variance is proportional to σ^2. For example, if a number of subpopulations had different regression lines but shared a common σ, it would be appropriate, in obtaining a tolerance band for any one of the lines, to employ s based on the pooled observations. Alternatively, our technique would yield simultaneous tolerance bands for all the lines.

The set $S(x_0)$, regarded as a tolerance set corresponding to the single value x_0, has an inevitable redundancy because of the constraint that the intersection of all $U(x)$ must have probability at least δ; and the resulting tendency is to widen the tolerance interval. A "precise" calculation (subject to the approximation used) of the redundancy may be based on the fact that in (2) we are using the $1 - \delta$ point for χ_3^2, although the left member of the inequality is distributed as χ_2^2. If, for example, $\delta = .95$, so that $2c^2 = 7.815$, $P(U(\lambda, x))$ is just over .98.

TABLE 2

δ	γ	$\lvert x \rvert$	(a)	(b)	(c)	(d)	(e)
.99	.95	0	3.61	7.79	3.97	4.79	5.75
.99	.95	0.393	6.58	7.94	7.23	5.65	6.15
.99	.95	0.430	6.99	7.97	7.70	5.77	6.19
.95	.75	0	1.77	3.59	2.13	2.67	2.78
.95	.75	0.393	3.23	3.78	3.89	3.34	2.97
.95	.75	0.430	3.43	3.82	4.14	3.43	3.05
.90	.50	0	0.95	1.91	1.33	1.71	1.38
.90	.50	0.393	1.73	2.18	2.42	2.30	1.69
.90	.50	0.430	1.84	2.22	2.58	2.39	1.76

Table 3 of Lieberman and Miller (1963) includes values of λ for three combinations of (δ, γ), and, for each combination, the same three values of x. The value of n was set at 15. In addition to their "Simultaneous Wallis" (a), they developed three other procedures termed by them "Simultaneous x and P central" (b), "Simultaneous x (fixed P) central" (c), and "Bonferroni" (d). Readers are referred to their paper for details. Table 2 (in which x has been standardized to conform with our notation) gives the values of λ obtained, that of the present paper being in the column (e).

It is reasonable to fear that $n = 15$ will make the approximation of Section 3 somewhat imprecise. If the result is a slightly large value for λ, we will not be unduly concerned; but we would not want λ to be underestimated as a result of $P(T)$ being less than δ. But an intuitive argument can be used to reassure us that if $P(T)$ is a bit less than δ, say $P(T) = \delta - \epsilon$, T can be enlarged to the desired size without λ being increased. We need only visualize attaching an increment $\Delta T(P(\Delta T) = \epsilon)$ to T, as much as possible of ΔT being at the bottom of the ellipsoid, the remainder at the top. Under the transformation $A = \bar{y} + bx$, $r = \lambda's$, each ellipse $U(x)$ will have an induced increment $\Delta U(x)$, the smallest set such that $\Delta T \subset U(x) + \Delta U(x)$. It is obvious that ΔT can always be so defined that $U(x) + \Delta U(x) \subset S(x)$ for all x.

5. Acknowledgment. I wish to express gratitude to Messrs. A. W. Douglas and D. M. Brown of Biometrics Research Services for valuable criticism of preliminary drafts of this paper.

REFERENCES

[1] LIEBERMAN, G. J. and MILLER, R. G., JR. (1963). Simultaneous tolerance intervals in regression. *Biometrika* **50** 155–68.

[2] WALD, A. and WOLFOWITZ, J. (1946). Tolerance limits for a normal distribution. *Ann. Math. Statist.* **17** 208–15.

[3] WALLIS, W. A. (1951). Tolerance limits for linear regressions. *Second Berkeley Symp. Math. Statist. Prob.* 43–51. Univ. of California Press.